B 80 28

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: METHOD FOR DETECTING TARGET SEQUENCES IN SMALL PROPORTIONS IN HETEROGENEOUS SAMPLES

(57) Abstract

The invention provides methods for detecting and analyzing molecules that exist in heterogeneous specimen or sample in small proportion relative to corresponding molecules.

METHOD FOR DETECTING TARGET SEQUENCES IN SMALL PROPORTIONS IN HETEROGENEOUS SAMPLES

Field of the Invention

The invention relates broadly to methods for detection and identification of nucleic acids that exist in a heterogeneous biological sample in low frequency.

Background of the Invention

5   It is often desirable to detect the presence in a complex biological sample of one or more molecules present in low frequency in the sample. For example, the detection of mutations in oncogenes at an early stage of oncogenesis are useful for early diagnosis of cancer. Such detection preferably is done in a specimen obtained through non-invasive, or minimally invasive means. Such specimens include stool, sputum, and other specimens that have a complex mixture of cellular components. DNA from cells

10   having mutations indicative of early-stage cancer are present in such specimens in low frequency with respect to wild-type DNA. Detection of a mutant DNA in the specimen using conventional techniques is often difficult because the specimen does not contain the DNA of interest, or the signal associated with such low-frequency DNA is

15   undetectable even if the target DNA is present in the specimen, or in a sample derived from the specimen. In contrast, disease-associated DNA is present in large amounts, and is easily detected in specimens, such as tumors, that are typically obtained by invasive means.

          With the advent of the polymerase chain reaction (PCR), detection of nucleic

20   acids became more routine, as the PCR allowed one to amplify vast quantities of a DNA of interest. Theoretically, PCR amplifies 100% of target, doubling the quantity of analyte with each cycle. Even with the abundance of material produced during PCR, careful attention must be paid to the amount of material presented to the PCR, and the representative nature of the input sample (that is, abnormalities must be sufficiently

25   represented in the input sample to assure detection). Practical PCR is not 100% efficient. In order to assure that PCR is being run with a reasonable level of specificity,

- 2 -

the T_m must be adjusted to reduce non-specific hybridization of primer. A consequence of increased specificity is a reduction in efficiency of the reaction. PCR becomes a stochastic process when it is not 100% efficient (i.e. a process subject to the laws of probability). Once the PCR reactants are in place (e.g., targets, sufficient primer,

5 polymerase, etc.), whether any specific target nucleic acid molecule is amplified is determined by the laws of probability.

For example, in a PCR having 30% efficiency (which is in the typical range for most PCRs), and in which 99 wild-type nucleic acids and 1 mutant nucleic acid are present in a sample obtained from a complex biological specimen (e.g., a sample, such

10 as stool, in which a target DNA is present in low frequency relative to other DNA, protein, etc. in the sample), there is nominally a 30% chance that the 1 mutant molecule will be amplified in the first round. If the mutant molecule is not amplified in the first round, its concentration in the sample will be reduced from 1 in 100 to about 1 in 130. If the mutant nucleic acid is not amplified in the first two rounds, it will exist in the sample

15 at an even lower ratio (about 1/169) with respect to the wild-type. Even if the mutant nucleic acid is amplified in every subsequent round of PCR in proportion to the wild type, its ratio in the sample will never be better than about 0.6% (1/169) of the sample (an approximately 40% reduction from its representation as compared to that before the amplification of two rounds). Thus, if an assay to detect the mutant nucleic acid in the

20 sample has a sensitivity limit for the mutant of 1%, it is unlikely that the mutant will be detected, even after amplification.

Similar problems exist in the detection of other low-frequency molecular species. For example, the detection of the relative amounts of high- and low- expression proteins may be undetectable over highly-expressed protein. A similar situation exists

25 in detecting RNA, and other cellular molecules. Accordingly, there is a need in the art for methods of detecting low-frequency molecular events, especially in heterogeneous biological samples. Such methods are presented by the invention, a brief description of which follows.

Summary of the Invention

30 Methods of the invention solve the problem of detecting low number, low-frequency molecular events in heterogeneous specimens. Methods of the invention

comprise determining the number of molecules in a sample that must be analyzed in order to maximize the probability that a low-frequency species will be detected in the sample. Methods of the invention are based upon a modeling of the stochastic effects in PCR. However, the principles disclosed herein are applicable to the identification,

5   detection and/or quantification of any low-frequency molecule, especially in a heterogeneous sample. Merely obtaining a large specimen (by weight or volume) is not sufficient if the specimen does not contain a sufficient number of target molecules, or if steps are not taken to assure that the minimum number of molecules are processed from the specimen into a sample to be analyzed.

10   The invention recognizes that there are two types of heterogeneity in a complex biological sample, such as stool, sputum, and others. A first type of heterogeneity is reflected in the relatively small amount of human DNA in such samples relative to other types of RNA and DNA (bacterial, viral, plant, and animal), proteins, etc. in the sample, and relative to other material such as mucus, fiber, etc. A second type of heterogeneity

15   is reflected in the relatively low amounts of a low-frequency human DNA (e.g., a mutant) with respect to the total human DNA in such samples. Thus, the detection of low-frequency human DNA (e.g., a mutant at the threshold of clinical relevance) is limited by the availability of such DNA in a sample prepared from a complex biological specimen.

20   Methods of the invention teach that the limited target DNA (corresponding, for example, to about 1% of the human DNA of a biological specimen) must be made available in a sample in order for amplification and detection to occur with high confidence. According to the invention, the number of molecules analyzed in a sample taken from a specimen determines the ability of the analysis to reliably detect low-

25   frequency DNA. In the case of PCR, the number of input molecules (mutant plus wild-type) must be about 500 or greater if the PCR efficiency is close to 100%, the low-frequency DNA exists as about 1% of the total sample DNA, and a 0.5% detection threshold is used. As PCR efficiency goes down, the required number of input molecules goes up. Analyzing the minimum number of input molecules determined to

30   be necessary by methods of the invention reduces the probability that a low-frequency event is not detected in PCR because it is not presented to the PCR or is not amplified

- 4 -

in the first few rounds. Methods of the invention comprise determining a threshold
number of sample molecules that must be analyzed in order to detect a low-frequency
molecular event at a prescribed level of confidence. Methods of the invention also
address the threshold number of molecules necessary for detection of a low-frequency

5   species given a predetermined level of assay sensitivity.

In a preferred embodiment, methods of the invention are applied to PCR
analysis. Stochastic errors in the diagnosis of a mutant nucleic acid result from a failure
to present sufficient relevant nucleic acid to the PCR, or from the failure to amplify a
relevant nucleic acid. Using a model of stochastic errors in PCR, the invention

10   provides a method for determining the minimum number of molecules that must be
analyzed in order to provide confidence that: 1) the detection of signal associated with a
low-frequency molecule is indicative of the actual presence in the sample of that
molecule, and is not due to background "noise"; and 2) that the absence of signal is
indicative of the absence of the target molecule, and not a failure to detect the low-

15   frequency molecule

Practical (i.e., non-theoretical) PCR is not a noise-free amplifier. In any PCR that
is not 100% efficient there is some level of stochastic noise (failure to amplify a target
DNA due to failure to prime template). In order to reduce the level of noise due to non-
specific primer binding, primer hybridization conditions typically are set so that as little

20   non-specific binding as possible occurs. However, the higher the specificity of primer
hybridization, the lower, necessarily, the efficiency of the PCR. Thus, in order to
assure appropriate specificity, PCR efficiency is usually between about 2% and about
40%, especially when working with highly heterogeneous samples like stool, sputum,
cervical scrapings, etc. Greater PCR efficiencies are routinely achieved when

25   amplifying, for example, plasmid DNA which does not have the heterogeneity of
samples used for human diagnostics and screening. According to the invention, PCR
at those efficiencies inevitably introduces stochastic errors when a target for
amplification is in low frequency in the sample due to a failure to prime the low
frequency DNA.

30   A PCR efficiency of 30% means that, in any one round of PCR, 30% of the target
will be amplified, producing about 1.3X molecules as compared to the previous round

- 5 -

(assume that PCR primers are placed outside the region of mutation, and amplify through the mutation). If the number of mutant molecules is high as, for example, in a tumor specimen, mutant DNA will almost certainly be amplified. It is only in the case of a heterogeneous sample in which the mutant DNA exists in small proportion that

5   stochastic effects described herein play a role in reducing the probability of amplifying the mutant DNA. However, a typical cancer-associated mutant DNA in the early stages of oncogenesis represents about 1% of the DNA in a heterogeneous sample. If PCR efficiency is set at 30% because of constraints needed to assure specific amplification, each mutant DNA molecule has only a 30% chance of being amplified in any round of

10   PCR. If no mutant is amplified in the first round, the mutant DNA will represent only about 0.7% of the DNA in the sample after round 1. If no mutant is amplified in the first two rounds (0.7 x 0.7, or a 49% probability), the mutant DNA will represent about 0.6% of the DNA in the sample going into round three of the PCR . If the post-amplification assay used to detect the mutant has a sensitivity of no more than 0.5% for the mutant,

15   it may not be possible to reliably detect the presence of the mutant. This is not the case when a mutant DNA species is present in large amounts relative to the wild-type DNA in a specimen (e.g., in a tumor) because there will be numerically sufficient mutant material in any prepared sample, thereby increasing the likelihood of target amplification. Also this is not the case when analyzing a heterogeneous sample when a

20   great deal of material is present. Intuitively, 10,000,000 total input molecules. If 1% is mutant then 100,000 mutants exist. 100,000 molecules will be more or less faithfully amplified in early rounds of PCR (even at low efficiency) in a way that may not be the case for 1 or 2 mutant molecules. Methods of the invention are also applied to the detection and analysis of infectious organisms. (e.g., the presence of minimum residual

25   disease (for example, HIV) in blood)

        The problems associated with detecting low-frequency molecules have been overcome by methods of the invention which provide means for determining the threshold number of molecules that must be involved in, for example, a PCR, in order to assure, within a predefined degree of statistical confidence, that low-frequency

30   molecules actually are detected. Methods of the invention are used to determine the minimum number of molecules that must be analyzed to assure detection of a low-

frequency sample molecule in any assay system or systems in which stochastic processes operate. However, for ease of exemplification, methods of the invention will be provided in the context of conducting PCR in a heterogeneous sample. Once the minimum number of molecules that must be analyzed is determined, the skilled artisan

5 can use any method available in the art to prepare (from a biological specimen) a sample having at least that number of molecules. One method, exemplified herein, is to homogenize the specimen, or a portion thereof, in a physiologically-compatible buffer at a volume (ml) to mass (mg) ratio of at least 5:1, and preferably about 10:1 or 20:1 and extract DNA therefrom. Sample dilution assists in releasing DNA from the complex

10 elements present in a heterogeneous sample, and is one way in which to ensure that the number of mutants is sufficient for detection. In a preferred method, the sample is enriched for human DNA using techniques known in the art such as sequence-specific capture prior to amplification. Other methods for increasing overall DNA (e.g., total human DNA) are also applicable for use in methods of the invention.

15 In general, methods of the invention comprise detecting and/or quantifying a target nucleic acid in a biological sample such as, for example, tissue or body fluids. Methods of the invention may be practiced by preparing a sample comprising a minimum number of nucleic acid molecules sufficient to detect a target nucleic acid and then detecting said target nucleic acid and/or quantifying the number of target nucleic

20 acid molecules in a sample. In a preferred method, the target nucleic acid is amplified prior to the step of detecting/quantifying the target nucleic acid.

In preferred methods of detecting and/or quantifying a target nucleic acid, the target nucleic acid is a low-frequency molecule such as a mutant nucleic acid. In a highly preferred embodiment, the target nucleic acid is present in said sample at about

25 between 0.5% and about 10% of the total species-specific nucleic acid in the sample.

Methods of the invention further comprise amplifying a target nucleic acid know or suspected to be present in a biological specimen. In one embodiment, a method of amplifying a target nucleic acid comprises preparing a sample comprising a minimum number of target nucleic acids present in said sample at about between 0.5% and

30 about 10% of the total species-specific nucleic acid in the sample and amplifying the target nucleic acid. The method may further comprise the step of detecting the

amplified target nucleic acid.

An alternative method for amplifying a mutant nucleic acid comprises selecting an amplification efficiency, level of statistical confidence, and suspected ratio of nucleic acid having a mutation to total nucleic acid in said specimen; determining a minimum
5   number of nucleic acid molecules that must enter an amplification reaction in order to assure that a mutant nucleic acid will be amplified at a defined level of statistical confidence; preparing a sample comprising the minimum number of molecules sufficient to detect a mutant nucleic acid; and amplifying the mutant nucleic acid. In a preferred embodiment, the mutant nucleic acid is amplified by PCR.

10   The present invention also provides methods for detecting loss of heterozygosity in nucleic acid molecules in a biological specimen. Methods of the invention comprise preparing a sample comprising a minimum number of nucleic acid molecules necessary to detect a loss of heterozygosity, enumerating a number of target nucleic acid molecules suspected of having a loss of heterozygosity and a reference number of non-
15   target nucleic acid molecules, and comparing the target number to the reference number. Methods of the invention determine whether the difference between the number of target and reference nucleic acid molecules is statistically significant, a statistically significant difference being indicative of a loss in heterozygosity.

According to preferred embodiments of the invention, any method for identifying
20   low-frequency molecules may be employed . In a preferred embodiment, the low frequency molecules are amplified by, for example, PCR prior to detecting the low-frequency molecules. Examples of preferred methods include those disclosed in U.S. Patent No. 5,670,325, incorporated by reference herein. A highly-preferred post-amplification detection means is the use of single-base extension assays to detect
25   and/or identify a single nucleotide at, for example, a polymorphic locus.

Methods of the invention may be performed on any biological specimen. Methods of the invention are most advantageous when performed on a heterogeneous sample such as tissue and body fluid in which the detection is desired of a molecule that is present in the sample in small amount relative to other molecules in the sample.
30   A stool sample is a good example of a heterogeneous sample in which a mutant DNA, for example a mutant oncogene or tumor suppressor, is present at very low levels

- 8 -

relative to other nucleic acids in the sample at early stages of oncogenesis. Diagnosis of such mutant DNA at early stages in the development of, for example, colorectal cancer is advantageous because colorectal cancer is highly-curable if detected at early stages. Methods of the invention provide means to increase the likelihood of detection of mutant DNA indicative of the early stages of disease, such as cancer. Particularly preferred biological specimens include blood, biopsy tissue, sputum, pus, semen, saliva, stool, lymph, cerebrospinal fluid, and urine.

5

Methods of the invention are also useful in the detection of a low-frequency molecule in specimens, especially heterogeneous tissue or body fluid specimens, obtained by pooling samples from multiple individuals or from identified populations (e.g. healthy, diseased, heterozygotes, etc.). Pooled samples may be used to identify clinically-relevant loci (e.g., single nucleotide variants associated with disease or pharmacological efficacy, safety, etc.), or to screen numerous patients simultaneously for a mutation. DNA isolated from pooled specimens or samples may also be used.

10

An example of the use of methods of the invention is provided below. The skilled artisan recognizes that the principles of the invention are applicable to a wide range of assays, including amplification reactions, competitive hybridizations, and other assays in which a low-frequency molecule is detected in a heterogeneous specimen or sample. The inventive methods are provided in the context of PCR for exemplification and illustration of a preferred embodiment for practice of the methods.

15

20

Description of the Drawings

Figure 1A is a flow chart of a model program for determining the minimum number of molecules that must be analyzed to assure detection of low frequency molecules in a heterogeneous sample.

Figure 1B is a flow chart of the stochastic PCR sampling routine shown as "Take Stochastic Sample of Mutant to be Presented to PCR" in Figure 1A.

25

Figure 1C is a flow chart of the stochastic PCR routine shown as "Perform stochastic PCR cycle" in Figure 1A.

Detailed Description of the Invention

The invention provides methods for determining the minimum number of

30

molecules that must be analyzed in order to provide statistical confidence that a low-frequency molecule or molecular event will be detected in a sample prepared from a biological specimen, especially a heterogeneous biological specimen. Methods of the invention capitalize on the realization that a sample containing a minimum number of

5   molecules overcomes stochastic sampling errors. Identifying a minimum threshold of molecules for analysis assures, within a defined level of statistical confidence, that a low-frequency molecule is detected if it is present.

In the context of the PCR, methods of the invention, as described below, provide statistical confidence that a low-frequency DNA will be amplified in at least the early

10   rounds of PCR, thereby preserving the ratio of that DNA with respect to the total molecules in the sample – even after further rounds of PCR amplification. Thus, methods of the invention are especially useful when primers for PCR are designed to hybridize with template in a region outside the suspected mutation. The primers will be extended through the region of mutation, thus producing amplicon that corresponds to

15   either the wild-type sequence or to the mutant sequence (depending on which template the primer anneals to). If the mutant sequence exists in low proportion relative to the wild-type, and if PCR is run at below 100% efficiency, stochastic effects begin to take over, as primer may anneal to the wild-type nucleic acid more frequently than to the nucleic acid containing a mutation in the region to be amplified. Methods of the

20   invention are also useful if primers are designed to hybridize in the region of a mutation that differs from wild-type by only one or two bases, and annealing stringency is such that the mutant-directed probes non-specifically hybridize with the wild-type sequence.

Exemplification of the invention is based upon a model of stochastic processes in PCR. The model operates by iterating stochastic processes over a number of PCRs.

25   The model incorporates a preset PCR efficiency (established to meet separate specificity requirements), and a preset ratio of mutant DNA to total DNA in the sample to be analyzed (which is a property of the disease to be detected and the nature of the sample. For example, in stool samples, it is thought that a >1% ratio of mutant DNA to total human DNA is associated with disease.). Based upon those input values, the

30   model predicts the number of molecules that must be presented to the PCR in order to ensure, within a predefined level of statistical confidence, that a low-frequency molecule

- 10 -

will be amplified and detected. Once the number of molecules is determined, the skilled artisan can determine the sample size to be used (e.g, the weight, volume, etc.), depending on the characteristics of the sample (e.g., its source, molecular makeup, etc.).

5   I.   **Model of Stochastic Processes in PCR**

A model used to exemplify methods of the invention is presented. Other methods and models are available to the skilled artisan, and can be used to implement the invention to determine the minimum number of molecules necessary to detect a low-frequency DNA. The model according to methods of the invention solves the
10   problems associated with amplification of low-frequency DNA. The model dictates the number of molecules that must be presented to the PCR in order to reliably ensure amplification and detection.

The exemplary model simulates selection of DNA for amplification through several rounds of PCR. For purposes of the model, a sample is chosen that contains a
15   ratio of mutant-to-total DNA of 1:100, which is assumed to lie at the clinical threshold for disease. For example, in colorectal cancer 1% of the human DNA in a specimen (e.g., stool) is mutated (i.e., has a deletion, substitution, rearrangement, inversion, or other sequence that is different than a corresponding wild-type sequence). Over a large number of PCR rounds, both the mutant and wild-type molecules will be selected (i.e.,
20   amplified) according to their ratio in the specimen (here, nominally 1 in 100), assuming there are any abnormal molecules in the sample. However, in any one round, the number of each species that is amplified is determined according to a Poisson distribution. Over many rounds, the process is subject to stochastic errors that, as described above, reduce the ability to detect low-frequency mutant DNA. However, the
25   earlier rounds of PCR (principally, the first two rounds) are proportionately more important when a low-frequency species is to be detected (for the reasons discussed above),and any rounds after round 10 are virtually unimportant. Thus, the model determines the combined probability of (1) sufficient mutant molecules being presented to the PCR, and (2) the effects of stochastic amplification on those molecules so that at
30   the output of the PCR there will be a sufficient number of molecules and a sufficient ratio of mutant to total molecules to assure reliable detection

- 11 -

The model used to run the number of molecules necessary at the first round of PCR was generated as a "Monte Carlo" simulation of a thousand experiments, each experiment consisting of 10 cycles of PCR operating on each molecule in the sample. The simulation analyzed (1) taking a sample from the specimen; and (2) each round of
5 PCR iteratively to determine whether, for each round, a mutant DNA if present in the sample was amplified. Upon completion of the iterative sampling, the model determined the percent of rounds in which a mutant strand was amplified, the percent of mutants exceeding a predetermined threshold for detection (in this example 0.5% based upon the mutant:total ratio of 1%), the coefficient of variation (CV) for stochastic
10 sampling in each round alone, and the coefficient of variance for stochastic sampling and PCR in combination.

Stochastic noise is created in PCR if the PCR efficiency is anything other than 0% or 100% (these two cases represent either there is no amplification at all or perfect fidelity of specific amplification). The noise, or background, signal level in a PCR that is
15 between 0% and 100% varies with the efficiency of the PCR. The standard deviation of stochastic noise, S, in a PCR is given by the equation, $S = \sqrt{npg}$, where n is the number of molecules in the sample, p is the efficiency of PCR, and q is 1-p. Table 1 presents results obtained for iterative samplings with PCR efficiency set at 100% and 20%, and a mutant:total ratio of 0.5%.

20 Table 1 represents output from the model in 12 experiments conducted under various conditions. The first row shows the nominal number of molecules entering the first round of PCR (i.e., the total number of molecules available for amplification). The second row shows the percent of molecules (DNA) in the biological specimen that is expected to be mutant. For colorectal cancer indicia in DNA recovered from stool, the
25 threshold for clinical relevance in the detection of early stage cancer is 1%. That is, 1% of the DNA in a sample derived from a heterogeneous specimen (e.g., stool) contains a mutation associated with colorectal cancer. The 6th row is the threshold of detection of the assay used to measure PCR product after completion of PCR. That number is significant, as will be seen below, because sufficient mutant DNA must be produced by
30 PCR to be detectable over aberrant signal from wild-type and random background noise. Under the heading "Outputs", the first line provides the likelihood that at least

- 12 -

one mutant molecule is presented to the first round of PCR. The second line under the Output heading provides the likelihood of detection of mutants (after PCR) above the predetermined threshold for detection. For example, in experiment 4, the results indicate that in 87.9% of experiments run under the conditions specified for experiment 4, the number of mutants will exceed the threshold number for detection. Finally, the last two rows provide the coefficient of variation for sampling, and for the combination of sampling and PCR.

TABLE 1

	100% Efficiency PCR						20% Efficiency PCR					
Inputs	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8	Exp 9	Exp 10	Exp 11	Exp 12
Nominal number of Molecules going into PCR	50	100	200	500	1,000	10,000	50	100	200	500	1,000	10,000
Percent of molecules that are mutant	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Number of PCR Rounds	10	10	10	10	10	10	10	10	10	10	10	10
Efficiency of PCR per round	100%	100%	100%	100%	100%	100%	20%	20%	20%	20%	20%	20%
Number of experiments modeled	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
M.W ratio which can be reliably detected by assay	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%
Outputs												
% of Experiments each with the number of mutant cells >0	39.30%	64.40%	87.30%	99.40%	100%	100%	39.10%	63.50%	86.10%	99.60%	100%	100%
% of Experiments each with number of mutant cells exceeding threshold	39.30%	64.40%	87.30%	87.90%	87.10%	100%	38.20%	52.40%	63.30%	83.30%	91.60%	100%
Stochastic Sampling CV	143.20%	100%	70%	44.10%	32.10%	9.90%	144.60%	101.80%	72.00%	44.20%	32.00%	10%
Stochastic Sampling and Stochastic PCR CV	143.20%	100%	70%	44.10%	32.10%	9.90%	179.70%	123.10%	92.10%	86.40%	40.30%	12.70%

As shown in Table 1, even at 100% PCR efficiency, mutant DNA is detected in only 97.1% of the samples when 1000 input molecules are used (i.e., 1000 DNA molecules are available for priming at the initial PCR cycle), even though 100% of the DNA is amplified in any given round of PCR. When 10,000 molecules are presented, it is virtually certain that the mutant DNA will be amplified and detected, as shown in the results for experiment 6 in Table 1. Stochastic errors due to variation in the number of input molecules become less significant at about 500 input molecules and higher (i.e., the CV for stochastic variations is about the same regardless of whether PCR efficiency is 20% or 100%). At lower PCR efficiency (20% in Table 1), the model shows that introducing 50, 100, 200, 500, or even 1000 molecules into the PCR does not assure either amplification or detection. As shown in experiment 12, introducing 10,000 molecules results in amplification of the mutant target, and a high likelihood of its subsequent detection. Thus, even with 100% efficient PCR, significant false negative events occur when input molecules fall below 500.

- 13 -

The foregoing analysis shows that there is a unique range for the number of molecules that must be presented to a PCR in order to achieve amplification of a low-frequency DNA, and to allow its detection. That range is a function of the PCR efficiency, and the percentage of low-frequency (mutant) DNA in the sample, and the

5 detection threshold. The aforementioned model was developed and run in Visual Basic for Applications code (Microsoft, Office 97) to simulate a PCR as described above. A flow chart containing the programming steps is provided in Figure 1. The statistical confidence level within which results were measured was held constant at approximately 99%. Only the PCR efficiency and percent mutant DNA were varied. As

10 discussed above, the model iteratively samples DNA in a "Monte Carlo" simulation over a thousand experiments, each experiment consisting of 10 rounds of PCR. The results are shown below in Table II.

TABLE II

PCR Efficiency	Number of molecules needed			
	1% Mutant	2% Mutant	5% Mutant	10% Mutant
10%	3,000			
20%	2,500			
50%	2,200			
100%	1,600			
10%		1,500		
20%		1,200		
50%		1,000		
100%		800		
10%			500	
20%			450	
50%			400	
100%			300	
10%				225
20%				200
50%				150
100%				125

Regression of the data obtained using the model as described above, produced the set of curves set forth below in Table III.

5

TABLE III

Molecules needed to overcome stochastic effects with about 99% confidence
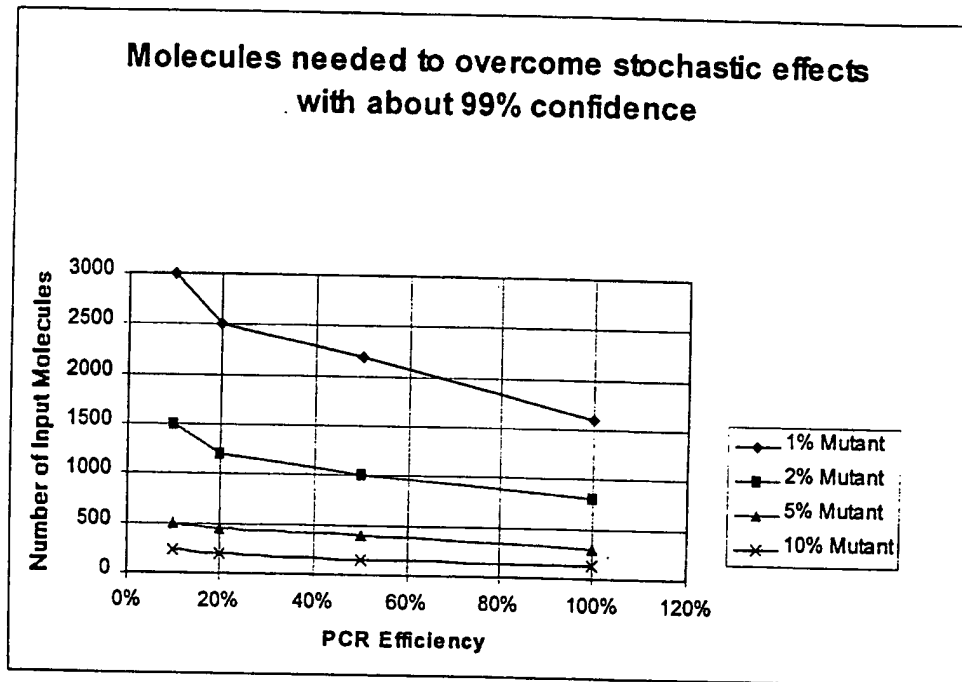
Using Table III, the optimal number of molecules to be presented to the PCR is
determined by selecting a PCR efficiency (or determining the efficiency by empirical
means), and selecting a percentage of the sample suspected to be mutant DNA
associated with disease. This, in turn, dictates a threshold of detection. Not all
detection strategies have similar underlying detection thresholds, so an appropriate
technology must be selected. The percentage mutant DNA may be determined by
clinical considerations as outlined above for colorectal cancer.

In practice of the invention, one may determine the PCR efficiency and percent
expected mutant in order to maximize the probability of obtaining amplified, detectable
mutant DNA. For example, one may select N, the number of input molecules from the
"1%" curve in Table III, when 5% of the sample is expected to be mutant DNA in order
to increase the confidence of the assay result.

Once the number of molecules for input to the PCR is determined, a sample
comprising that number of molecules (or greater) is prepared for PCR according to
standard methods. The number of molecules in a sample may be determined directly
by, for example, enumerative methods such as those taught in U.S. Patent No. 5,

- 16 -

670,325, incorporated by reference herein. Alternatively, the number of molecules in a complex sample may be determined by molar concentration, molecular weight, or by other means known in the art. The amount of DNA in a sample may be determined by mass spectrometry, optical density, or other means known in the art. The number of

5   molecules in a sample derived from a biological specimen may be determined by numerous means in the art, including those disclosed in U.S. Patent Nos. 5,741,650 and 5,670,325, both of which are incorporated by reference herein.

In one preferred embodiment, a sample is prepared from a stool specimen by homogenizing in a physiologically-compatible buffer at a stool mass to buffer volume

10   ratio of about 20:1 in order to maximize the amount of DNA in the sample available for amplification. Physiologically acceptable buffers include those solvents generally known to those skilled in the art as suitable for dispersion of biological sample material. Such solvents include phosphate-buffered saline comprising a salt, such as 20-100mM NaCl or KCl, and optionally a detergent, such as 1-10% SDS or Triton™, and/or a

15   proteinase, such as proteinase K (at, e.g., about 20mg/ml). A preferred solvent is a physiologically-compatible buffer comprising, for example, 1M Tris, 0.5M EDTA, 5M NaCl and water to a final concentration of 500 mM Tris, 16mM EDTA and 10mM NaCl at pH 9. The buffer acts as a solvent to disperse the solid stool sample during homogenization and to facilitate separation of the DNA from the bacterial and fibrous

20   components. Increasing the volume of solvent in relation to solid mass of the sample results in increased yields of DNA.

Buffer is added to the solid sample in a solvent volume to solid mass ratio of at least about 5:1. The solvent volume to solid mass ratio is preferably in the range of about 10:1 to about 30:1, and more preferably in the range of about 10:1 to about 20:1.

25   Most preferably, the solvent volume to solid mass ratio is about 10:1. Typically, solvent volume may be measured in milliliters, and solid mass measured in milligrams, but the practitioner will appreciate that the ratio of volume to mass remains constant, regardless of scale up or down of the particular mass and volume units. That is, solvent volume to solid mass ratios may be measured as liters:grams or μl: μg. The

30   minimum number of DNA molecules in the prepared sample may be verified by molarity, optical density, enumeration, or other means known in the art.

- 17 -

After PCR amplification, assays are performed to detect the presence of mutant DNA in the amplified sample. Such mutant DNA may be detected in enumerative methods (see above) or by bulk detection using, for example, fluorescent markers, mass markers, radioactive markers, and the like. Once methods of the invention are used to ensure that low-frequency material, if present, will be amplified for detection, the means for measuring the presence in the amplified sample of the low-frequency DNA is immaterial to the invention. Such means may be chosen by the skilled artisan in accordance with available materials, convenience, and clinical or diagnostic requirements.

Claims

1   1.   A method for detecting a target nucleic acid known or suspected to be present in
2   a biological specimen, the method comprising the steps of:
3   preparing a sample comprising a minimum number of nucleic acid molecules
4   sufficient to detect a target nucleic acid; and
5   detecting said target nucleic acid in said sample.

1   2.   The method of claim 1, wherein said biological specimen is a tissue or body fluid.

1   3.   The method of claim 1, wherein said target nucleic acid is a mutant nucleic acid.

1   4.   The method of claim 1, wherein said target nucleic acid is present in said sample
2   at about between 0.5% and about 10% of the total species-specific nucleic acid in said
3   sample.

1   5.   The method of claim 1, further comprising the step of amplifying said target
2   nucleic acid prior to detecting said target nucleic acid.

1   6   A method for quantifying the amount of a target nucleic acid in a biological
2   specimen, the method comprising the steps of:
3   preparing a sample comprising a minimum number of nucleic acid molecules
4   necessary to detect a target nucleic acid; and
5   enumerating the number of target nucleic acid molecules in said sample.

1   7.   A method for preparing a heterogeneous sample for detection of an analyte, the
2   method comprising the steps of:
3   determining a minimum number of analyte molecules that must be present in
4   said sample for detection of said analyte at a defined level of statistical confidence; and
5   preparing a sample comprising said minimum number of analyte molecules.

1   8.   A method for amplifying a target nucleic acid known or suspected to be present
2   in a biological specimen, the method comprising the steps of:
3   (a)   preparing a sample comprising a minimum number of molecules sufficient
4   for detection, within a defined degree of statistical confidence, of a target nucleic acid
5   present in said sample at between about 0.5% and about 10% of the total species-
6   specific nucleic acid in said sample; and
7   (b)   amplifying said target nucleic acid.

1   9.   The method of claim 8, further comprising the step of detecting amplified target

2 nucleic acid.

1 10.   The method of claim 8, wherein said biological specimen is a tissue or body fluid.

1 11.   The method of claim 8, wherein said specimen is stool.

1 12.   The method of claim 11, wherein said preparing step comprises homogenizing

2 said stool specimen in buffer at a stool sample mass-to-buffer volume ratio of about

3 20:1.

1 13.   The method of claim 11, wherein said preparing step comprises enriching said

2 specimen for human DNA.

1 14.   The method of claim 13, wherein said enriching step comprises sequence-

2 specific capture of human DNA.

1 15.   A method for amplifying a mutant nucleic acid in a sample prepared from a tissue

2 or body fluid specimen, comprising the steps of:

3   (a)   selecting an amplification efficiency, level of statistical confidence, and

4 suspected ratio of nucleic acid comprising a mutation to total nucleic acid in said

5 specimen;

6   (b)   determining, based upon said efficiency, said ratio, a minimum number of

7 nucleic acid molecules that must enter an amplification reaction in order to assure

8 within said level of statistical confidence, that a nucleic acid comprising said mutation

9 will be amplified;

10   (c)   preparing a sample comprising said minimum number of nucleic acid

11 molecules; and

12   (d)   amplifying a region of said nucleic acid suspected to contain said

13 mutation.

1 16.   The method of claim 15, wherein said amplifying step comprises a polymerase

2 chain reaction.

1 17.   A method for detecting loss of heterozygosity in nucleic acid molecules in a

2 biological specimen, the method comprising the steps of:

3   preparing a sample comprising a minimum number of nucleic acid molecules

4 necessary to detect a loss of heterozygosity;

5   enumerating a number of target nucleic acid molecules in said sample a subset

6 of which is suspected of having a loss of heterozygosity;

7   enumerating a reference number of non-target nucleic acid molecules in said

8   sample; and

9   comparing said target number to said reference number,

10   a statistically-significant difference between said target number and said reference

11   number being indicative of a loss of heterozygosity.

1   18.   The method of claims 1 or 8, wherein said biological specimen is obtained from a

2   pooled patient population.

1   19.   The method of claim 18 wherein said pooled biological specimen comprises a

2   stool sample obtained from members of a patient population.

1   20.   A method for detecting a mutant nucleic acid known or suspected to be present

2   in a biological specimen, the method comprising the steps of:

3   preparing a sample comprising a number of total nucleic acid copies sufficient to

4   detect a mutant nucleic acid with a predetermined level of statistical confidence if said

5   mutant nucleic acid is present in said sample; and

6   detecting said mutant nucleic acid in said sample.

INPUT:
- % of mutant in population
- efficiency of PCR
- # of molecules into PCR
- detection threshold
- # of PCR cycles
- # of experiments to be modeled

Take Stochastic sample of mutant to be presented to PCR (Figure 1B)

Perform Stochastic PCR cycle (Figure 1C)

No typically 10 cycles

PCR cycles = MAX ?

Yes

typically 1,000

No

Experiment = MAX?

Yes

Post Output
- 1/1 of experiment with number of mutants > 0
- 1/1 of experiment exceeding threshold
- CV of sampling
- CV of sampling & PCR

Figure 1A

Stochastic Sampling

```
                    ┌────────────────────┐
                    │ clear mutant counter│
                    └────────────────────┘
                              │
  ┌───────────────────────────┤
  │                           ▼
  │         ┌──────────────────────────────────┐
  │         │ • for each molecule, compare a    │
  │         │   random number to Mutant %.      │
  │         │ • If less than mutant % then      │
  │         │   bump mutant counter             │
  │         └──────────────────────────────────┘
  │                           │
  │                           ▼
  │              No       ╱ Mol =  ╲
  └──────────────────────▕  Mol MAX ▏
                          ╲         ╱
                              │
                              │ Yes
                              ▼
```

To PCR Cycle

Figure 1B

Sochastic PCR

inputs are: efficiency of PCR

| mutant counter = # of mutants | number of mutant & wild type |
| wild counter = # of wilds | molecules going into this round |

for each molecule

- for each mutant molecule:
  • Compare a random number to PCR eff.
  • If less thant PCR eff, bump mutant counter

- for each wild type molecule:
  • Compare a random number to PCR eff.
  • If less than PCR eff, bump wild type counter

Mutant
&
wild molecules
=
MAX?

No

Yes

• update # of mutant molecules
• update # of wild molecules

Figure 1C

	national Application No
	PCT/US 99/28064

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHEN M -S ET AL: "DETECTION OF SINGLE-BASE MUTATIONS BY A COMPETITIVE MOBILITY SHIFT ASSAY" ANALYTICAL BIOCHEMISTRY,US,ACADEMIC PRESS, SAN DIEGO, CA, vol. 239, no. 1, 15 July 1996 (1996-07-15), pages 61-69, XP000598275 ISSN: 0003-2697 see whole doc. esp. p.66, 2 .col, 4, par.	1-5,7,8
X	EP 0 648 845 A (EASTMAN KODAK CO) 19 April 1995 (1995-04-19) see whole doc. esp. claims and examples	1,7,8,15
X	US 5 506 105 A (HAYDOCK PAUL V) 9 April 1996 (1996-04-09) see claims, figures and abstract	1,2,4,5, 7,8,15

-/--

[X]	Further documents are listed in the continuation of box C.	[X]	Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 March 2000	20/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Müller, F

Form PCT/ISA/210 (second sheet) (July 1992)

International Application No

PCT/US 99/28064

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 23651 A (EXACT LAB INC) 3 July 1997 (1997-07-03) see whole doc. esp. claims and p.12,1.32 ff.	1-20
X,P	WO 99 07894 A (ALLEN PETER G ;KOLESAR JILL M (US); WISCONSIN ALUMNI RES FOUND (US) 18 February 1999 (1999-02-18) the whole document	1,6,7

1

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0648845	A	19-04-1995	JP	7163370 A	27-06-1995
			SG	68568 A	16-11-1999
			US	5674717 A	07-10-1997
US 5506105	A	09-04-1996	NONE		
WO 9723651	A	03-07-1997	US	5670325 A	23-09-1997
			AU	711754 B	21-10-1999
			AU	1430797 A	17-07-1997
			CA	2211702 A	03-07-1997
			EP	0815263 A	07-01-1998
			JP	10503384 T	31-03-1998
			US	6020137 A	01-02-2000
WO 9907894	A	18-02-1999	US	6013442 A	11-01-2000
			AU	8693798 A	01-03-1999